BBC RD 1974/37

1969 1974
THE QUEEN'S AWARD TO INDUSTRY

BBC

RESEARCH DEPARTMENT

REPORT

DIGITAL VIDEO: some bit-rate reduction methods which preserve information in broadcast-quality digital video signals

D.F. Reid, B.Sc.

Research Department, Engineering Division
THE BRITISH BROADCASTING CORPORATION October 1974

BBC RD 1974/37
UDC 621.376.56
621.391.
621.397.

DIGITAL VIDEO: SOME BIT-RATE REDUCTION METHODS WHICH PRESERVE INFORMATION IN BROADCAST-QUALITY DIGITAL VIDEO SIGNALS

D.F. Reid, B.Sc.

Summary

PCM digital video signals require a very high bit rate and for economic reasons, particularly in transmission, it is desirable to minimise the required bit rate. Methods for reducing the bit rate fall into two main categories: (a) those methods which remove information to which the human eye is insensitive, and (b) those which exploit the redundancy of the picture signal. In (a), the original signal cannot be reconstructed whereas in (b) it can, i.e. the method is described as being reversible in that all the information contained in the original signal can be retrieved.

For further investigations into reversible bit-rate reduction methods it is useful to consider the entropy of a digital signal. This gives the minimum bit rate necessary to transmit all the information contained in the signal, and its value depends on the signal statistics. One measurement technique, described in detail, is used to determine entropy of 8 bits/sample p.c.m. PAL signals derived from four colour television still pictures. It is estimated that the information content of most broadcast quality 8 bits/sample p.c.m. signals range from about 3½ bits/sample to about 6½ bits/sample. This gives some indication of the lowest bit rates which could be produced by a reversible coder.

Some reversible coding techniques reduce the bit rate by assigning short code words to frequent sample values and longer code words to less frequent sample values. This has the disadvantages that the required instrumentation can be very complex, and that the bit stream is non-uniform and requires a buffer store to produce bits at a uniform rate. The picture may also be more seriously affected by transmission errors than with other coding techniques. An alternative scheme is studied in detail, and it is concluded that this reversible system could be adapted to operate for p.c.m. signals or differential p.c.m. (d.p.c.m.) signals, and in the case of 8-bit p.c.m. signals, might give a reduction of about two bits/sample.

Issued under the authority of

P. Rainger

Head of Research Department

Research Department, Engineering Division,
BRITISH BROADCASTING CORPORATION

October 1974

EL-97

DIGITAL VIDEO: SOME BIT-RATE REDUCTION METHODS WHICH PRESERVE INFORMATION IN BROADCAST-QUALITY DIGITAL VIDEO SIGNALS

DIGITAL VIDEO: SOME BIT-RATE REDUCTION METHODS WHICH PRESERVE INFORMATION IN BROADCAST-QUALITY DIGITAL VIDEO SIGNALS

D.F. Reid, B.Sc.

1. Introduction

This report describes an investigation to determine the feasibility of reducing the bit rate of digital video signals in such a way that the original pulse-code modulated (p.c.m.) digital signal can be recovered exactly. Such a process is said to be reversible and may be applied to any digital signal, not necessarily a p.c.m. signal.

Popular methods for reducing the bit rate of digital video signals are based on transform coding[1] and differential pulse-code modulation (d.p.c.m.).[2,3] These methods tend to rely on the fact that certain types of picture information or detail are such that the human eye is insensitive to their being reproduced with low accuracy. When these methods are applied to a p.c.m. coded signal, 'rounding-off' errors and other slight inaccuracies are introduced so that the original p.c.m. signal cannot be recovered exactly, i.e. the process is non-reversible.

Reversible bit-rate reduction can be achieved by assigning short code words to very probable events and longer code words to improbable events. This constitutes a variable-length code.[4] A variable-length code may be designed to match any one source using very simple procedures. One such procedure was devised by Shannon and Fano[5] and another by Huffman.[6] Fig. 1 shows how reversible bit-rate reduction could be applied, allowing for processes such as error correcting and transmission coding to be involved in making digital source signals suitable for transmission. Variable-length codes produce bits at a non-uniform rate. A buffer store[7,8,9,10] is therefore required to give a constant bit rate for transmission.

PCM coding of video signals is, in general, wasteful of bits since equal numbers of bits are used to describe equal areas of picture irrespective of whether the areas contain large amounts of fine detail or plain regions.

The information contained in a p.c.m. signal may be well below that which could be transmitted by the bit-rate employed, i.e. there is some redundancy. The aim of a reversible bit-rate reduction system must be to reduce the redundancy while leaving the information unaltered.[11]

Although the investigation described in this report is mainly theoretical some measurements have been made on broadcast-quality digital PAL video signals derived from slides to determine how much information they contained. The variation in information content of the pictures derived from different slides was used as a measure of the degree of adaptability which would be required of a reversible system.

This report discusses the theory which allows us to postulate a reversible system, and investigates a practical reversible coder.

2. Theory

2.1. Information theory and entropy

2.1.1. Information source

An information source[12,13,14,15] can be regarded as a 'black box' producing a word, or a message consisting of a series of words, to be communicated to the receiving terminal. An information source may be said to have a source alphabet which is the collection of all words which that source emits.

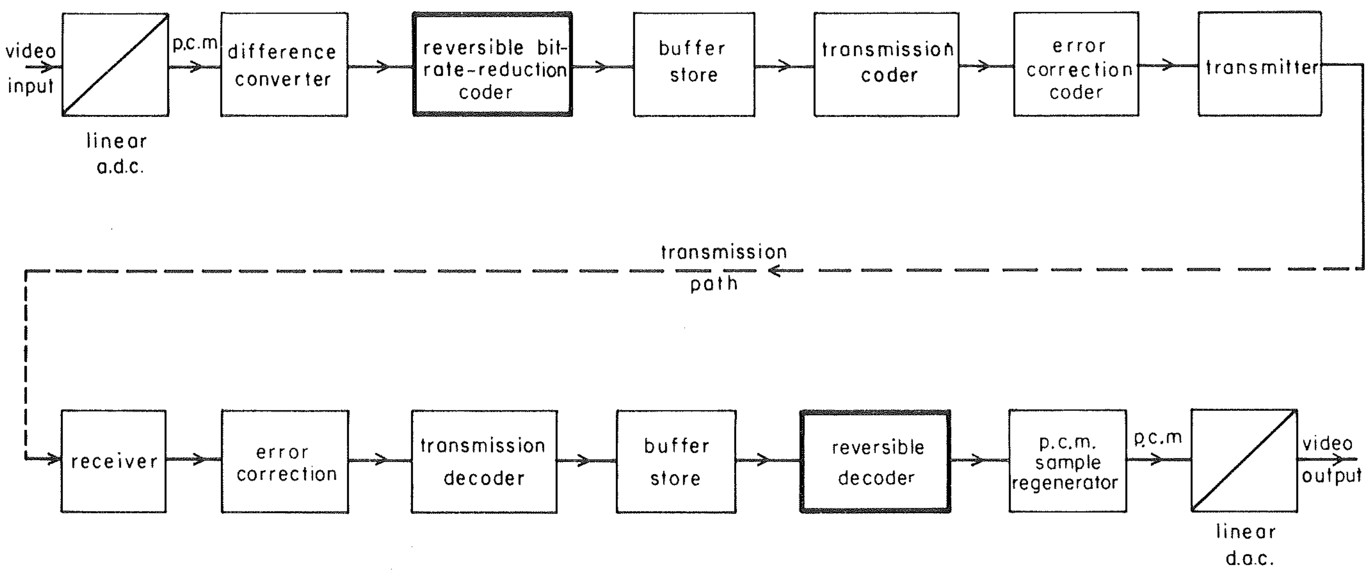


Fig. 1 - A digital video transmission system using variable-length coding

A probability-density function (p.d.f.) is associated with an information source and specifies the probability of occurrence of each word of the alphabet. These probabilities may be conditional, i.e. their values may depend on some previous output(s). For sources in which the p.d.f. changes with time the source statistics are said to be non-stationary. If successive words are statistically independent the source is said to be a zero-memory, or memoryless, source. If the occurrence of a source word depends on a finite number, m, of preceding words then the source is said to be an mth order Markov source. A memoryless source is a zero-order Markov source.

2.1.2. Entropy and information

In information theory, the information content[5,15,16] of a message or word from a source is defined as the negative of the logarithm of the probability 'p' that this message or word will be emitted from this source, i.e.

$$\text{Information content} = -\log p. \qquad (1)$$

If the base of the logarithm is 2, then the units for information content are bits. Thus, a word with a probability of ½ has an information content of one bit.

The *average* information content per word emitted from a source is often referred to as the entropy*[5] of the signal and is usually denoted by the letter H. In mathematical terms,

$$H \text{ in bits} = -\sum_i p_i \log_2 p_i. \qquad (2)$$

where p_i is the probability of word i and the summation includes all possible words.

The usefulness of entropy, expressed in bits, is that it gives the minimum average number of 'yes' or 'no' answers (or bits) required per word to define sequences of words which include all possible words in proportion to their probability of occurrence.

As an example of entropy calculations, consider p.c.m. coding in which the values of regularly spaced samples of an analogue signal are quantised using 256 possible levels, and each level is represented by a separate 8-digit binary number, i.e. the code gives 8 bits per sample. If each level is equally probable, then p_i = 1/256 for all levels and therefore the entropy of the signal is given by

$$H = -256\ (1/256)\ \log_2\ (1/256) = 8 \text{ bits}$$

This result, that the entropy is equal to the actual number of bits per sample, indicates that the p.c.m. code uses the minimum possible number of bits per sample when all levels are equally probable. It will be found, however, that if all levels are not equally probable, then the entropy will be less than 8 bits per sample. For example, if 128 of the levels have a probability of 1/128 and the remaining levels have zero probability, then the entropy equals 7 bits, indicating that in this case all the information in the 8-bit p.c.m. code could be transmitted with only 7 bits per sample.

Although the term 'picture entropy' has gained widespread acceptance the term is misleading because it usually refers to the entropy of a digital video signal, which depends on more than the picture properties. Apart from the original scene, the entropy of this signal depends on

(i) The scanning method (e.g. a 625-line signal has more information than a 405-line signal).

(ii) The signal bandwidth.

(iii) The type of digital coding employed.

(iv) Any processing carried out on the digital signal[17] (e.g. taking differences between adjacent samples of a p.c.m. signal takes account of correlation and gives a signal with a lower entropy than that of the p.c.m. signal).

2.1.3. Redundancy and correlation

It is wasteful of communications capacity to allocate a constant number of bits to a message describing several events which are not equiprobable. The redundancy of a p.c.m. signal, in bits/sample, is the difference between the number of bits/sample transmitted and the entropy of the signal. This difference is representative of the possible saving in the number of bits required to describe the events. Redundancy is a property of the signal produced from the output of an information source, not of the information source alone; if one bit is added to each message describing the output of a source, the redundancy of the signal is increased by 1 bit/sample whereas the information content is unaltered.

It is helpful to consider two types of redundancy which may exist within a signal produced by an information source. The first is statistical redundancy; equal numbers of bits are used to describe events which are not equiprobable. The second is correlation redundancy; a degree of correlation exists between successive 'events' or samples. The effect of this may be examined by considering not simply the p.d.f. of the samples, but various so-called conditional p.d.f.s relating to the probabilities of certain combinations of samples. A simple example is the p.d.f. of the

* In thermodynamics and physics the term entropy denotes disorder whereas in communication engineering entropy is a measure of information. There is no contradiction here. From Equation (1), in communication engineering, unlikely events convey greatest amounts of information. If we are very uncertain about which event will occur, a large amount of information is conveyed when we know the outcome. A television picture with much detail may have a waveform close to that of random noise. A picture of random noise displays complete disorder. Each possible brightness level is of low probability (but equiprobable) so Equation (2) has a maximum value and by our definition, the signal describing the disordered picture conveys greatest information.

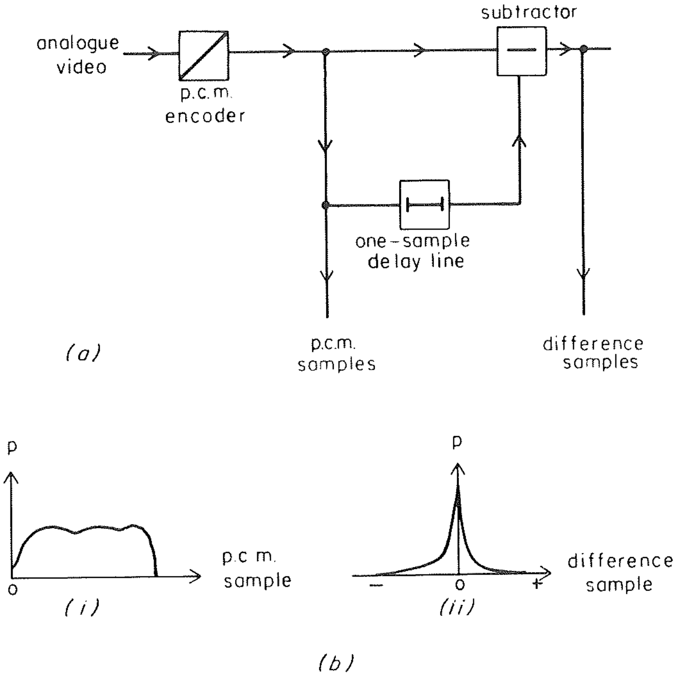


Fig. 2
(*a*) A differential p.c.m. encoder
(*b*) Probability density functions
(i) PCM samples (ii) difference samples

differences between adjacent samples. An entropy figure based on a particular conditional p.d.f. of the source (the conditional entropy) therefore gives a target for the minimum bit-rate necessary to describe the source output if all the correlation redundancy shown by that conditional p.d.f. were removed. For all sources with correlation this figure would be less than the entropy calculated by ignoring such correlation. To ignore all correlation is to treat each sample independently of all others. Entropy figures calculated from p.d.f.s obtained by ignoring all correlation will be called zero-order entropies. Zero-order entropy is the minimum bit-rate necessary to transmit the source information when successive samples are coded independently.

Various means can be employed to exploit the correlation redundancy within a signal. For example, in video p.c.m., differences between adjacent samples can be taken as shown in Fig. 2, to form a difference signal. Because this operation is reversible the total conditional entropies of both the p.c.m. signal and the difference signal are equal but the zero-order entropy of the *difference* signal is generally lower than that of the p.c.m. signal, indicating that a saving could be achieved by coding successive differences independently.

2.1.4. Adaptivity

A coder which adapts itself to changes in source statistics is said to be adaptive. An adaptive coder could be designed to adapt to changes which occur from one picture to another or from one area of a picture to another area. It can be shown that an adaptive coder can reduce the required bit-rate below that of a non-adaptive-coder. (See Appendix 1.)

Entropy is a lower bound on the average number of bits required to describe the information produced by a source. However, values of entropy are dependent on probability distributions and for a video signal, these change as the picture changes and are also different for different areas of a given picture. In short, video information sources are not ergodic.*[5]

2.2. Reversible coding for minimum bit rate

2.2.1. General

Reversible bit-rate reducing coders generally employ a variable-length code in which the code words are not of uniform length. Bit-rate reduction will occur if the shortest code words are assigned to the most probable events, etc.[18]

Morse code is a well known example of a variable-length code. The dots and dashes are analogous to '1's and '0's in a digital system. The most common letter in the alphabets of several languages (E) is given one symbol while less common letters (e.g. Q, Z) are given four symbols. In the same way that 8-bit p.c.m. signals require framing (synchronisation) bits, Morse code requires a pause between letters and a longer pause between words to make it reversible. Otherwise, the code for A(·—) could be received as E(·) T(—) and the words FOR and GET could be received as FORGET, etc.

A code in which no code word is a prefix** of any other code word is said to be uniquely decipherable,[19,20,21] i.e. no message sequence will be ambiguous. As a result, framing information is not required (although it may be advantageous) for a uniquely decipherable code.

A near-optimum uniquely decipherable code in which

$$H \leqslant \bar{L} \leqslant (H + 1) \qquad (3)$$

where $\bar{L}$ is the average number of bits per sample, may be obtained for any probability distribution.***[22] Simple procedures for designing such reversible codes were devised by Shannon, Fano and Huffman.

2.2.2. Shannon-Fano codes

These are codes which have been designed[22] using the procedure developed independently by both Shannon and Fano.[5]

* Ergodic signals are statistically homogeneous. Word probabilities are obtained by measuring word frequencies in a particular sequence. The word probabilities from an ergodic source will approach definite values as the length of the sequence is increased and these probabilities will be independent of the particular sequence chosen. Video information sources are ergodic for sequencies which are very long compared to one picture, but for short sequencies they are not ergodic.

** i.e. no complete code word is equivalent to the beginning of another code word.

*** Actually $H \leqslant \bar{L} \leqslant H + 1 + N2^{1-M} \log_2 (N-H)$
where N is the number of code words and M is the maximum permissible length of code word. If no constraint is imposed on M, i.e. $M \to \infty$, this relation reduces to relation (3).

As an example of a reversible Shannon-Fano code consider the following. Imagine an information source with eight possible outputs, whose probabilities p_i are as shown in column 1, of Table 1, listed in decreasing value. Note that $\sum_i p_i = 1$. The probabilities are separated into two groups, so that the sum of the probabilities in one group is approximately equal to the sum of the probabilities in the other group, by drawing a horizontal line. By convention, a 0 is assigned to all probabilities above the line and a 1 is assigned to all probabilities below the line. The sub-division process is repeated until each group contains only one number. The length of a code word appropriate to each probability is given by L_i, the number of times that probability was involved in a sub-division, and is shown in column 3. The Shannon-Fano code for each input is given by the 0's and 1's for each sub-division and is shown in column 4.

If this code were used the average length of codeword would be

$$L = \sum_i p_i L_i = 2{\cdot}80 \text{ bits}$$

whereas a uniform code (one whose code words are of constant length) would use 3 bits per sample to code each of the eight events. It is interesting to calculate the entropy of the probability distribution in Table 1.

$$H = -\sum_i p_i \log_2 p_i = 2{\cdot}706 \text{ bits/sample}$$

Thus Relation (3), Section 2.2.1, is satisfied.

TABLE 1

Derivation of a Shannon-Fano Code

1	2				3	4
p_i	1	2	3	4	L_i	Code
.36		0			2	00
	0					
.12		1			2	01
.12			0		3	100
		0				
.12			1		3	101
	1					
.07				0	4	1100
			0			
.07				1	4	1101
		1				
.07				0	4	1110
			1			
.07				1	4	1111

TABLE 2

Derivation of a Huffman Code

p_i	1	2	3	4	5	6	7	L_i	Code
							1		
						.62			
					.38	.38			
.36	.36		.36	.36	.36			2	00
				.26	.26				
			.24	.24					
		.14	.14	.14					
	.14	.14	.14						
.12	.12	.12	.12					3	011
.12	.12	.12						3	100
.12	.12	.12						3	101
.07	.07							3	110
.07	.07							3	111
.07								4	0100
.07								4	0101

2.2.3. Huffman codes

These are codes which have been designed using the procedure developed by Huffman.[6]

An example of a reversible Huffman code is derived in Table 2 for the information source discussed in Section 2.2.2. Step 1 involves adding together the two smallest probabilities to form a larger probability which is inserted into the list according to its size. The process is repeated until the set has only one member, whose value is 1 since $\sum_i p_i = 1$.

The length of a code word for the event whose probability is x is given by the number of times that probability is involved in a summation. By convention, if the probability is uppermost in the pair being added together, a zero is assigned to that step. The code for the event with probability x is obtained by tracing out the path followed by x, working back from probability 1 to x, using the rules given above. The average length of a code word for the Huffman code derived in Table 2 is given by

$$L = \sum_i p_i L_i = 2{\cdot}78 \text{ bits and } H = 2{\cdot}706 \text{ bits/sample,}$$

as before.

Hence Equation (3), Section 2.2.1, is again satisfied.

The following points should be noted:

(a) Shannon-Fano and Huffman codes for the same probability distribution are different.

(b) Both codes give similar average code-word lengths.

(c) The average code-word lengths are both greater than the entropy figure.

(d) Both codes are uniquely decipherable. No code word is a prefix of any other code word.

(e) If a probability distribution changes very significantly the Shannon-Fano or Huffman code may be far from optimum for the new distribution.

2.2.4. Run-length codes

A run-length code[24] is a code in which a sequence of identical symbols (or words) is replaced by a code indicating both the length of the sequence and the symbol (or words) of which is consists. The aim of a run-length code is to reduce the number of bits required for coding, without loss of information. To achieve this, run-length codes are usually variable-length codes. Run-length codes are useful where the number of code words is small and/or long runs are highly probable.

2.2.5. Buffer stores

Variable-length codes produce bits at a varying rate. In order to minimise the bandwidth required for the transmission of digital signals, digits must be transmitted at a constant rate. This is achieved by feeding bits into a buffer store as they are generated and reading out at a constant rate.[7,8,9,10]

The process is analogous to pouring water into a bucket in an irregular way, yet having a constant flow of water out of a tap in the bottom of the bucket. If the water comes in too quickly to too slowly the bucket may overflow or it may run dry (underflow).

When underflow occurs in a buffer store dummy bits are sent. Although this is wasteful of bits, no information is lost. No matter how large the store is made, there will always be a finite probability of overflow. The probability of overflow can be greatly reduced if the read-out rate is made 2 or 3% higher than the average rate at which bits are produced.[7] Remedial action will be necessary on overflow if breaks in the transmission of information are to be avoided. For example, the store could be completely or partially emptied and an alternative coding system could send the missing samples. A suitable standby system for digital video transmission would be a system which gave acceptable quality and used fewer bits per sample than were being transmitted per sample before overflow.

3. Entropy measurements of video information sources

3.1. General

A reversible coder will produce at least as many bits per sample, on average, as the entropy of the signal which is being coded. Reversible coders can be quite complex and would not be justified if considerable savings in bit rate could not be achieved, or if non-reversible systems were acceptable. Measurements were therefore made to estimate how much information is contained in typical still television pictures.

3.2. Probability-density function measurements

To calculate the entropy of an information source we must know the probability-density function of the source alphabet. Probability-density functions were obtained for the information sources derived from the four colour slides shown in Fig. 3, as follows.

The video signals were first PAL-coded and then quantised to 8-bit accuracy at a sampling rate equal to three times the colour subcarrier frequency. Exact differences were taken, using nine-bit arithmetic to avoid rounding errors, between samples three sample periods apart using the apparatus shown in Fig. 2. These differences were treated as the output of a memoryless information source. The probabilities of all the different possible values of this difference signal were obtained, using a digital counter to record the number of times a given difference occurred during a period of one second.

The probability-density functions for 'Girl with head-scarf', 'Bowl of fruit' and 'Crowd scene' are shown in Fig. 4

(*a*) Girl with headscarf

(*b*) Crowd scene

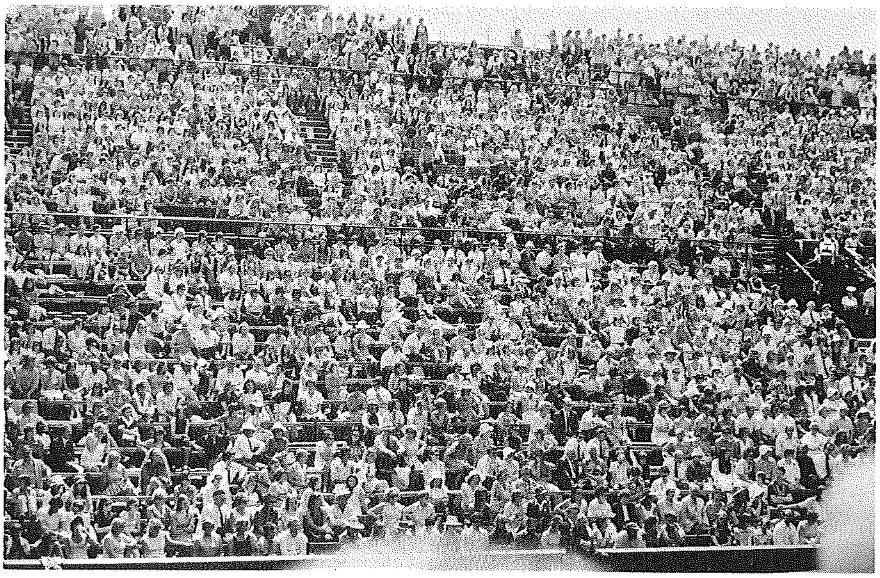
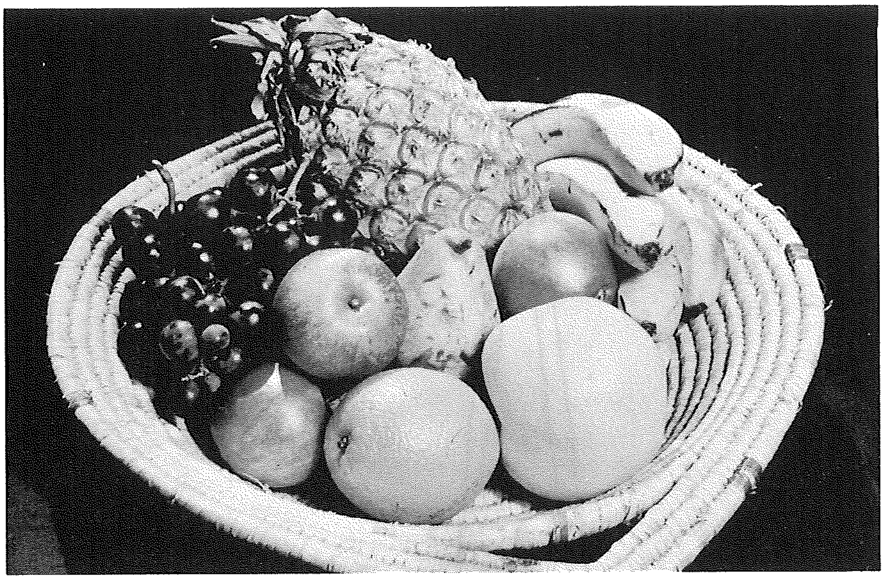


Fig. 3 - Colour slides which were used for measurements

for difference values between +24 and −24. 'Girl with sunflowers' has been omitted as it is very similar to 'Bowl of fruit'. Usually $P_{|i|} \geqslant P_{|i|+1}$ and such a distribution will be called 'typical'. Fig. 4 shows that all pictures tested had approximately typical probability-density functions. The probability-density functions apply to the picture information only. Samples occurring during the line and field blanking periods of the video signal have been ignored.

The probability-density functions in Fig. 4 were used to obtain the probability-density functions of the 5-bit per sample signal which would be obtained if the 9-bit difference signal were passed through the non-linear quantiser used in a particular 5-bit/sample d.p.c.m. coder. This non-linear quantiser reduced the number of bits per sample by assigning single codes to groups of adjacent values of the 9-bit difference signal (Appendix of Reference 2). Difference-sample probabilities were grouped in this way to form the p.d.f. histograms shown in Fig. 5.

Fig. 8 (Section 7.2) shows that the difference samples plotted in Fig. 4 follow Cauchy statistics. The entropy of such difference signals may be calculated from the probability of the difference-sample $i = 0$. This property could prove to be very useful in an adaptive reversible coder. If a measure of the probability of occurrence of difference-



(*c*) Bowl of fruit

(*d*) Girl with sunflowers

Fig. 3 - Colour slides which were used for measurements

sample i = 0 is available then there will be a running indication of the minimum number of bits per sample which the coder could produce.

3.3. Entropy calculations

An entropy figure may be calculated for any probability-density function. Such an entropy figure would be the amount of information produced by an information source whose source alphabet had that probability-density function. The entropies of signals produced from information sources with Cauchy and Laplacian probability density functions (See Appendix, Section 7.2) were calculated using a programmable calculator and are shown in Fig. 6. These entropies are plotted as a function of the probability of the most common difference sample, i = 0.

Entropies were calculated using Equation (2), for the probability-density functions relating to the differences between samples one subcarrier period apart, for the pictures shown in Fig. 3. The results are shown in column 2 of Table 3. The resulting minimum bit-rates attributable to these pictures are shown in column 3. These figures were obtained by multiplying the figures in column 2 by $1{\cdot}33 \times 10^7$, this being the number of samples per second (3 x colour subcarrier frequency). Entropies were also calculated for the probability-density functions which would have arisen if line and field blanking intervals were ignored

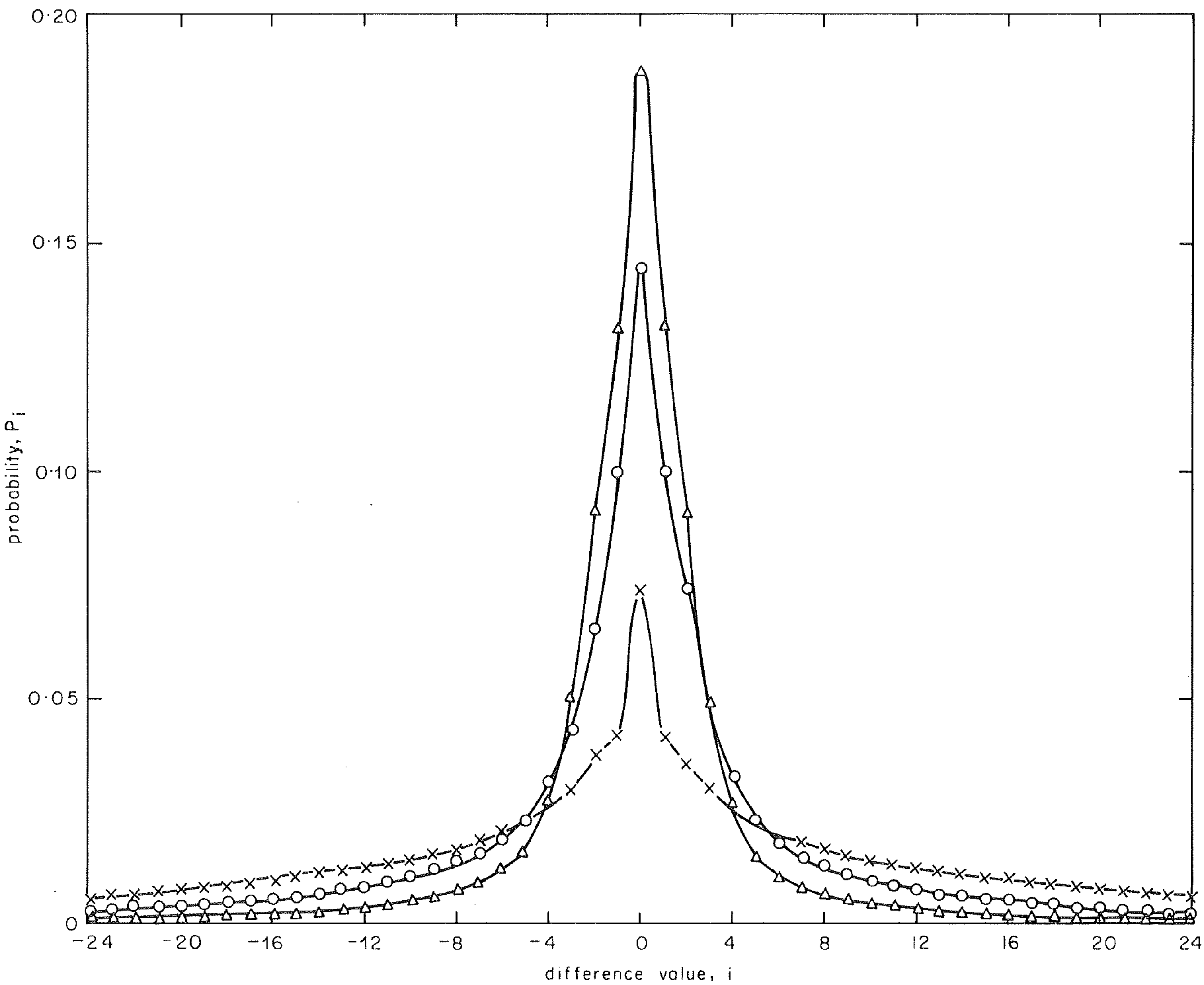


Fig. 4 - Probability density functions for 9-bit difference samples
△ Girl with headscarf X Crowd scene O Bowl of fruit

These p.d.f.'s are shown in Fig. 4 and refer to approximately $1{\cdot}02 \times 10^7$ video samples per second. The entropies are shown in column 4 of Table 3. The resulting minimum bit rates attributable to the video information are shown in column 5 of Table 3. These figures were obtained by multiplying the figures in column 4 by $1{\cdot}02 \times 10^7$. The figures in column 6 are the zero-order entropies of the information sources produced by a 5 bits/sample d.p.c.m. coder.[2] These entropy figures should be compared with 5 bits/sample as this is the number of bits used to transmit these d.p.c.m. signals. The entropy of the 5 bits/sample d.p.c.m. signal for 'Crowd scene' is 4·46 bits/sample, indicating that there is very little redundancy in this particular signal. This fact is verified by the comparatively flat probability-density function for 'Crowd scene' in Fig. 5.

4. Practical reversible systems

4.1. General

Difficulties might arise in the instrumentation of a reversible variable-length coder for digital video signals. Also, such a coder could suffer from severe limitations in practice, e.g. it might only cater for 'average' pictures. Section 4.2 describes a practical system that has been proposed and Section 4.3 discusses reversible coding for broadcast-quality digital video signals.

Exact differences between 8-bit samples spaced one subcarrier period apart, for broadcast colour pictures, have probability-density functions similar to those shown in Fig. 4. In general difference samples from an 8-bit p.c.m. signal could have 511 values (−255 to +255 inclusive) but in System I colour television these difference samples can have only approximate 320 values. A variable-length code for a source producing these samples would therefore have about 320 different code words and no code word should be a prefix of any other (see Section 2.2.1). Generating these code words requires a complicated coder. If the code (e.g. a Shannon-Fano or Huffman code) were designed for an 'average' picture it would not be optimum for very detailed (active) or very plain (inactive) pictures. Nevertheless, it would not be entirely unsuitable unless picture statistics became very 'non-typical' (see Section 3.2). A code could, however, be designed to cope with changing probability-density functions. Gilbert[22] discusses this problem and suggests two solutions. One is to limit the

TABLE 3

Zero-order Difference Sample Entropies and Bit Rates

1	2	3	4	5	6
Picture	9-bit difference signals				5-bit d.p.c.m. signals video only entropy, bits/sample
	video + syncs		video only		
	entropy bits/sample	bit rate Mb/s	entropy bits/sample	bit rate Mb/s	
(a) Girl with headscarf	3·65	48·6	4·21	43·0	3·65
(b) Crowd scene	5·03	67·0	5·80	59·2	4·46
(c) Bowl of fruit	4·21	56·0	4·87	49·6	4·00
(d) Girl with sunflowers	–	–	4·83	49·2	–

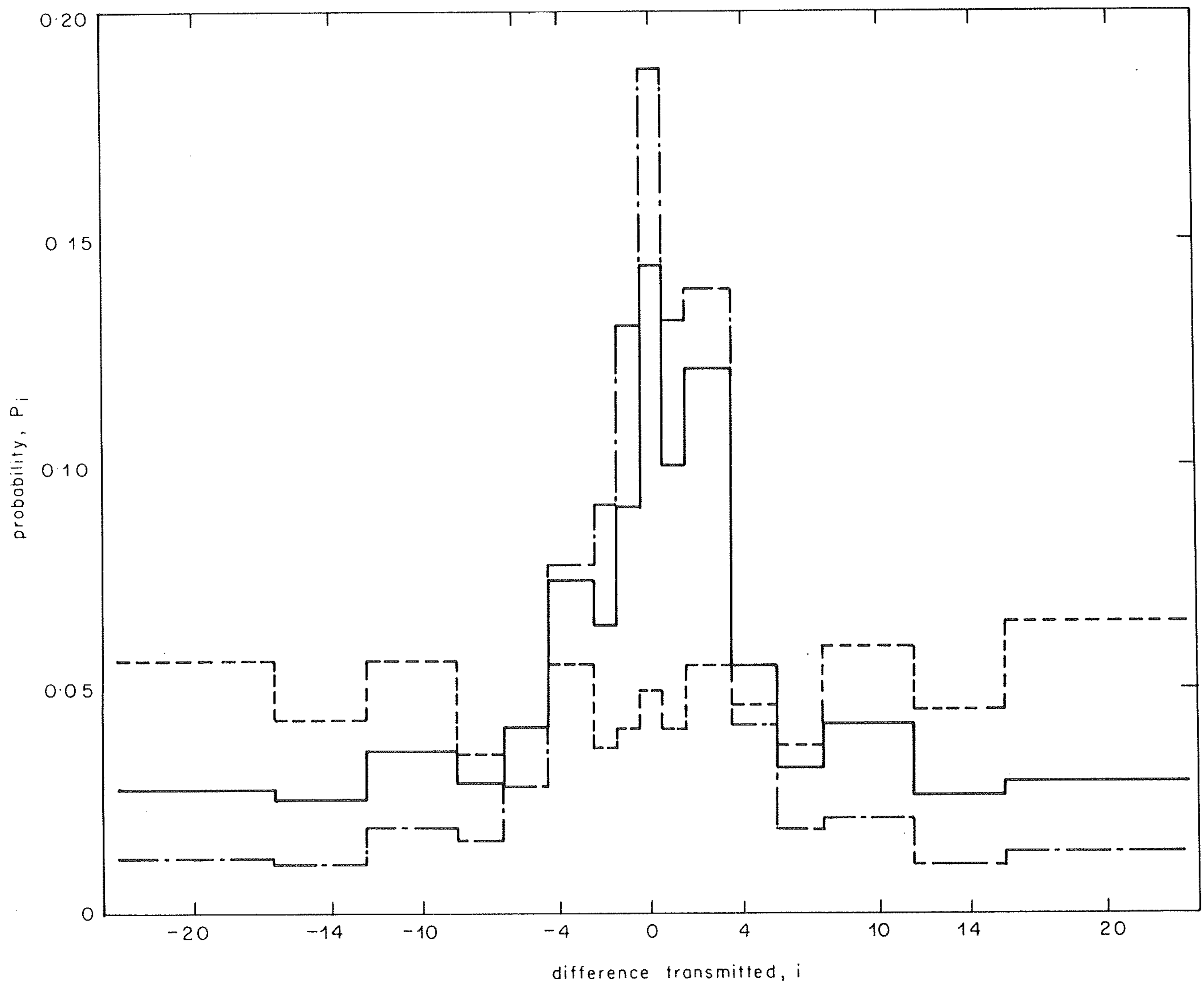


Fig. 5 - Probability density histograms for 5-bit tapered d.p.c.m. law

Bowl of fruit — Crowd scene — Girl with headscarf

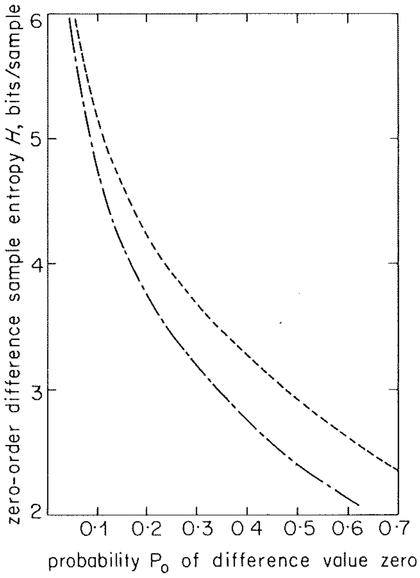


Fig. 6 - Entropies of signals produced from sources with Cauchy and Laplacian probability-density functions
——— Laplace distribution - - - - - - Cauchy distribution

maximum number of digits per code word and to minimise the average length of code word ($\sum_i p_i L_i$) subject to this constraint. If the picture changes and long code words become more probable the effect will not be so severe. Another approach is to design a code which would be as efficient as possible for two sets of signal statistics. Gilbert suggests that to simplify design procedure one set of statistics should have symbol probabilities equal.

Another way of dealing with varying signal statistics is to employ an adaptive code. An adaptive coder, designed by Rice and Plaunt, is discussed in the next Section.

The amount of buffer storage required by a reversible coder is best determined by simulation and subjective evaluation.

4.2. The Rice Machine

4.2.1. General

This Section discusses a coder which is information-preserving yet uses a very simple variable-length code. The Rice Machine[25,26] is the name given to an adaptive, reversible coder developed by Robert F. Rice. This coder was designed to accept pictures of widely-varying entropies in the planned 'Grand Tour' spacecraft mission to the outer planets, and it is expected to produce an average number of bits/sample which never differs from the entropy by more than 0·3 bits/sample. The three main constraints imposed upon the design of this coder were that it should be information preserving, that it should reduce the bit rate for very widely-varying pictures and that it should use only one-dimensional correlation. A coder satisfying these requirements is also of interest for reducing the bit-rate of broadcast digital video signals. However, the Rice Machine, unlike coding systems required for broadcast video signals, does not operate in 'real time', but it may be possible to use similar principles for operating on such signals. Picture information is stored in a large store after bit-rate reduction has occurred and is transmitted at a later time and at a very slow rate. Data compression, as described below, allows more pictures to be stored.

The coder operates on 8 bits/sample p.c.m. video signals and consists of a 'basic compressor' and a 'split-sample mode selector'. The mode selector decides, on a line-to-line basis, how many bits of each sample should be processed by the basic compressor. The basic compressor codes the p.c.m. samples by taking differences between adjacent samples and reducing the bit rate using the best one of four possible methods. The method may be changed at intervals of 21 consecutive samples, according to the data to be processed.

4.2.2. The basic compressor

The basic compressor takes differences between digital samples and, by taking account of sample-to-sample correlation, produces a typical distribution, or one which is very nearly typical (see Section 3.2). A block diagram of the basic compressor is shown in Fig. 7.

Exact difference samples are produced from n-bit* p.c.m. samples and these are coded using a comma code. A comma code is a geometric code[27] which can be very easily generated using a digital counter and a digital comparator, and is shown in Table 4. A comma coder addresses a '1' to an appropriate place in the buffer store. The output of the comma code generator is known as the fundamental sequence (f.s.). This fundamental sequence is coded by a variable-length code. The f.s. for each block of 21 eight-bit samples is treated as a string of triples, i.e. 3-bit words, so that there are now only eight possible code

* Sometimes the least significant bit(s) are almost random and are transmitted unchanged. The coder is said to be operating in a split-sample mode. This prevents the comma coder from generating very long words.

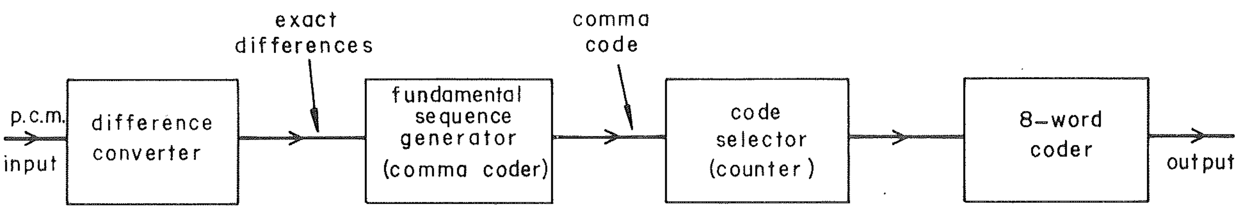


Fig. 7 - The basic compressor

TABLE 4

A Comma Code

Difference value d	Comma code
0	1
+1	0 1
−1	0 0 1
+2	0 0 0 1
−2	0 0 0 0 1
.	.
.	.
.	.

TABLE 5

8 Word Code

3-tuple	Code word
0 0 0	0
0 0 1	1 0 0
0 1 0	1 0 1
1 0 0	1 1 0
0 1 1	1 1 1 0 0
1 1 0	1 1 1 0 1
1 0 1	1 1 1 1 0
1 1 1	1 1 1 1 1

words; one or two bits may be added to permit this, and these triples are coded using an 8-word variable-length code. The 8-word code used in the Rice Machine is shown in Table 5. Note that it is uniquely decipherable. The 8-word code is invariant and will sometimes produce more bits than were contained in the input 8-bit p.c.m. signal. To counteract this and to use the 8-word code efficiently three options are provided and the one which results in the fewest number of bits is chosen. The decision levels may be related to the average length of the f.s. The three options and their decision criteria are as follows:

(a) If the f.s. is between 1 and 1·5 bits/sample the bit-inverted f.s. is fed through the 8-word coder.

(b) If the f.s. is between 1·5 and 3 bits/sample it is not coded, but is sent directly.

(c) If the f.s. is greater than 3 bits/sample it is fed through the 8-word coder.

Two bits are added to the f.s. corresponding to each block of 21 eight-bit samples to indicate which option has been chosen. The adaptability of the basic compressor comes from its ability to match the data to the code. No attempt is made to alter the variable-length code to suit the incoming data. Perhaps the 8-word code could be optimised for broadcast-type television signals. This would require an investigation into typical f.s.'s produced by broadcast picture signals.

The justification for the split-sample modes is that, due to camera noise and the analogue-to-digital conversion (a.d.c.) process, the least significant bit is almost completely random and is transmitted directly.

Therefore the entropy of the 8-bit signal is

$$H_8 \approx H_7 + 1$$

In fact, most 8-bit digital television signals are such[25,26] that

$$H_8 \approx H_6 + 2$$

and, quite frequently,

$$H_8 \approx H_5 + 3$$

Rice has not implemented an optimum variable-length code as one might be led to do after an examination of information theory. Instead he has coded sample differences with a deliberately non-optimum variable-length code (a comma code), treated the results as a source with a vocabulary of only eight words, and has coded these with what is effectively a run-length code of variable length. His method is adaptive but not in the usual sense. Data is processed to suit the code.

4.3. The suitability of reversible coding for broadcast pictures

What are the arguments for and against a reversible code for broadcast-quality pictures?

A transmission bit-rate reduction system must produce a constant output data rate. Most broadcast PAL digital video signals have zero-order entropy of the difference samples, of between 3½ and 6½ bits/sample. (See Table 3.) If the transmission rate is set at, say, 5 bits/sample, the buffer store will overflow for active pictures and as a result a standby system (see Section 2.2.5) will be frequently used if active pictures are being processed. If the standby system is used than complete reversibility is not obtainable. The variation in entropy for different pictures could be reduced in theory[28] by taking into account further correlation between samples. It is possible that an adaptive coder such as the Rice Machine will generate slightly fewer bits than the entropy figure in Table 3 and still be information-preserving. (See Appendix 1.) A non-adaptive coder would be suitable for picture statistics which are stationary with respect to time and picture area.

A reversible coder is well suited for non real-time systems because the above objections to the variations in entropy need not arise.

4.4. A comparison between d.p.c.m. and reversible coding

From theoretical considerations, O'Neal has shown[29] that entropy coding (i.e. coding with the aim of getting as close to the entropy figure as possible) of differences between p.c.m. samples could give up to 6 dB improvement in signal-to-quantising noise ratio compared with a d.p.c.m. coder employing an optimum non-linear quantising law and giving the same bit-rate as the entropy coder. However, the entropy coding system is difficult to implement and the full 6 dB improvement would be difficult to achieve with practical hardware.

DPCM is a successful bit-rate reduction method because for a considerable proportion of the time it reproduces a picture to the original quantising accuracy. Calculations based on the p.d.f. in Fig. 4 for 'Bowl of fruit' show that if this picture were coded using the 6 bits/sample coarse/fine d.p.c.m. law described in Reference 2, 8-bit quality would be obtained for 91% of the active video signal, and if it were coded using the 5 bits/sample d.p.c.m. law,* also described in Reference 2, 8-bit quality would be obtained for 68% of the active video signal.

A reversible coder only gives 8-bit quality if the output buffer store does not overflow. Thus, in practice, a reversible coder does not always give 8-bit quality. However, if, for a given size of buffer store, a reversible coder gives 8-bit quality for a significantly greater proportion of time than d.p.c.m. does, reversible coding offers distinct improvements in quality over d.p.c.m.

Assume that the output transmission rate (T) is fixed, at say, 5 bits/sample and that an average picture has an entropy (H) of 4·5 bits/sample. It was stated in Section 2.2.5 that P (the probability of overflow) is greatly reduced if T exceeds H by 2 to 3%. Here, T would exceed H by 11% for average pictures, and one could expect a buffer store of only moderate capacity to give a small value of P.

Thus, in principle, a reversible coder can be made to give better quality than d.p.c.m.

Table 3, column 6, implies that reversible coding for 5-bit d.p.c.m. signals could often reduce the bit-rate by approximately 1 bit/sample. As 5-bit d.p.c.m. has only 32 code words an adaptive variable-length code could be used instead of Rice's procedure.

5. Conclusions and recommendations

Relevant information theory has been discussed and reversible bit-rate reduction methods, based on entropy considerations, have been described. A practical investigation has shown that the zero-order entropy of exact differences between p.c.m. samples one subcarrier period apart for PAL video signals ranges from about 3½ bits/sample to 6½ bits/sample for broadcast-quality pictures coded to 8-bit accuracy. It is theoretically and instrumentally possible to build a coder which reduces the bit rate of digital video signals in a reversible way. However, such a coder would require a buffer store to provide the constant-rate output sequence required for transmission. Further work would be necessary to determine how closely the bit-rate in a practical reversible coder could approach the entropy of digital video signals, and how large the buffer store should be.

* Not necessarily recommended. Used here only as an illustration.

Reversible bit-rate reduction could be successfully applied to a digital signal which has redundancy. Taking differences and then coding reversibly would allow an 8-bit p.c.m. signal to be coded using from about 3½ to about 7 bits/sample depending on the picture. An 8-bit p.c.m. signal which has been reduced to 6 bits/sample by a d.p.c.m. process could be reduced further with no further degradation using a reversible coder. A 5-bit d.p.c.m. signal could, in some cases, be reduced reversibly to 4 bits/sample, but it has been demonstrated in this report that with some pictures no reversible bit-rate reduction would be possible.

Reversible bit-rate reduction could certainly be achieved in real time using fixed Shannon-Fano or Huffman codes, but in order to achieve a mean bit rate as low as the adaptive Rice coder it would probably be necessary to use an adaptive system with multiple coders, in which case the resulting complexity would probably be greater than that of the Rice coder.

Since a reversible coder would produce different numbers of bits for different pictures, reversible coding is more suitable for picture-storage systems than for real-time, fixed-rate transmission systems.

Further work would have to be done to determine how well a real-time reversible coder could be made to perform; this work would involve simulating a reversible coder to determine,

(a) the probability of overflow for various buffer store sizes,

(b) the acceptability of various standby systems,

(c) the difference between picture entropy and output bit rate,

for a variety of pictures.

Applied to broadcast quality p.c.m. video signals, reversible coding techniques appear to offer a reduction from eight to six bits/sample with virtually no impairment, but more cost and complexity, compared with six-bit d.p.c.m.

6. References

1. WINTZ, P.A. 1972. Transform picture coding. *Proc. IEEE*, 1972, **60**, 7, pp. 809 – 820.

2. DEVEREUX, V.G. 1973. Digital video: differential coding of PAL signals based on differences between

samples one subcarrier period apart. BBC Research Department Report No. 1973/7.

3. O'NEAL, J.B. 1965. Predictive quantising systems (differential pulse code modulation) for the transmission of television signals. *Bell Syst. Tech. J.*, 1966, **46**, pp. 689 – 722.

4. GILBERT, E.N. *and* MOORE, E.F. 1959. Variable length binary encodings. *Bell Syst. Tech. J.*, 1959, **XXXVIII**, 4, pp. 933 – 968.

5. SHANNON, C.E. 1948. A mathematical theory of communication. *Bell Syst. Tech. J.*, 1948, **27**, pp. 379 – 423 (Part I), pp. 623 – 656 (Part II).

6. HUFFMAN, D.A. 1952. A method for the construction of minimum-redundancy codes. *Proc. IRE*, 1952, pp. 1098 – 1101.

7. BUDRIKIS, HULLETT *and* PHIET. 1971. Transient-mode buffer stores for non-uniform code TV. *IEEE Trans. on Comm. Tech.*, 1971, **COM–19**, 6, pp. 913 – 922.

8. POPP, D.J. 1970. Buffer considerations for data compression of non-stationary video data. *IEEE Trans. on Comm. Tech.*, 1970, **COM–18**, 2, pp. 136 – 141.

9. SCHWARZ, G.R. 1968. Buffer design for data compression systems. *IEE Trans. on Comm. Tech.*, 1968, **COM–16**, 4, pp. 606 – 615.

10. DOR, N.M. 1967. Guide to the length of buffer storage required for random (Poisson) input and constant output rates. *IEEE Trans. on Electronic Computers*, Oct. 1967, pp. 683 – 684.

11. MUSSMAN, H.G. 1969. Redundancy reduction of p.c.m. signals. Symposium on Computer Processing in Communications, Polytechnic Inst. of Brooklyn, April 8–10, 1969. Wiley 1970.

12. ABRAMSON, N. 1963. Information theory and coding. McGraw-Hill, New York.

13. GALLAGER, R.G. 1968. Information theory and reliable communication. Wiley, New York.

14. BRILLOUIN, L. 1962. Science and information theory (second edition). New York, Academic Press.

15. ROTHSTEIN, J. 1951. Information, measurement and quantum mechanics. *Science*, Aug. 17, 1951, pp. 171 – 175.

16. ROTHSTEIN, J. 1952. Information and thermodynamics. *Phys. Rev. (letter)*, 1952, **85**, Jan. 1st, p. 135.

17. LIMB, J.O. 1968. Entropy of quantised television signals. *Proc. IEE*, 1968, **115**, 1, pp. 16 – 20.

18. TURNER, L.F. 1973. Data compression techniques as a means of reducing the storage requirements for satellite data – a quantitative comparison. Proc. of conf. on video and data recording, Birmingham, England, 10–12 July 1973 (London, England: IERE 1973) pp. 75 – 94.

19. SARDINAS *and* PATTERSON. 1953. A necessary and sufficient condition for the unique decomposition of coded messages. *IRE Conv. Record*, Pt. 8, pp. 104 – 108, 1953.

20. JAYNES, E.T. 1959. Note on unique decipherability, *IRE Trans. on Inf. Theory*, Sept. 1959, pp. 98 – 102.

21. McMILLAN, B. 1956. Two inequalities implied by unique decipherability. *IEEE Trans. on Inf. Theory*, Dec. 1959, pp. 115 – 116.

22. GILBERT, E.N. 1971. Codes based on inaccurate source probabilities. *IEE Trans. on Inf. Theory*, 1971, **IT–17**, 3, pp. 304 – 314.

23. CHOW, M.-C. 1970. Shannon-Fano encoding for differentially coded video telephone signals. Proc. Mervin J. Kelly coms. conf. Oct. 1970, Missouri.

24. GOLOMB, S.W. 1966. Run-length encodings. *IEE Trans. on Inf. Theory*, July 1966, pp. 399 – 401.

25. RICE, R.F. 1970. The Rice Machine: television data compression. Jet Propulsion Lab., C.I.T., California, Report No. 900-408, Sept. 1971.

26. RICE, R.F. *and* PLAUNT, J.R. 1971. Adaptive variable-length coding for efficient compression of spacecraft television data. *IEEE Trans. on Comm. Tech.*, 1971, **COM–19**, 6, pp. 889 – 897.

27. LIMB, J.O. 1970. Efficiency of variable length binary codes. Proc. Mervin J. Kelly, Comms. Conf., Oct. 1970, Missouri.

28. ISAILOVIC, J. 1973. Possible improvement in television signal prediction due to interframe correlation. *Electronics Letters*, 1973, **9**, 18, pp. 427 – 428.

29. O'NEAL, J.B. 1971. Entropy coding in speech and television differential p.c.m. systems. *IEEE Trans. on Inf. Theory*, Nov. 1971, pp. 758 – 761.

7. Appendix

7.1. Adaptive coding

Let the output of an information source, e.g. a digital video signal, be subdivided into blocks consisting of equal numbers of samples. These blocks could be pictures, areas of picture, or parts of one line.

Let the probability distribution of sample values in the kth block be $P[k]$. The entropy of the kth block is $H(P[k])$.

The average entropy of N blocks is therefore

$$\hat{H} = \frac{1}{N} \sum_{k=1}^{N} H(P[k])$$

Let the composite probability distribution for N blocks be

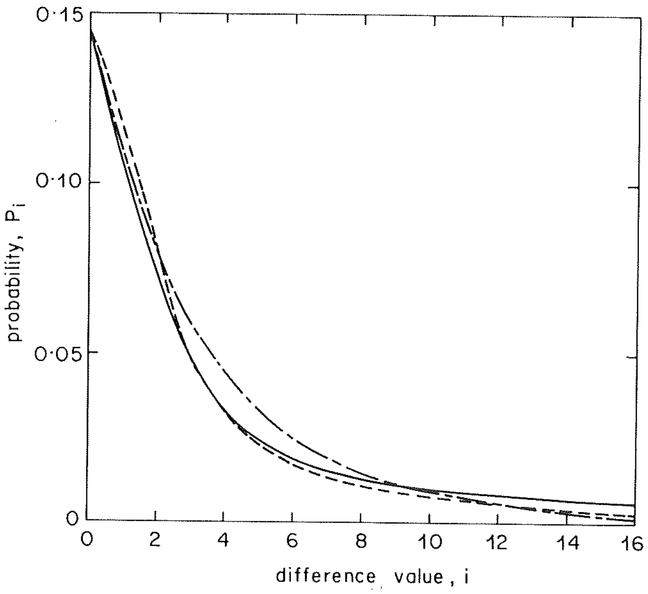
$$P^* = \frac{1}{N} \sum_{k=1}^{N} P[k]$$

The entropy of this distribution is

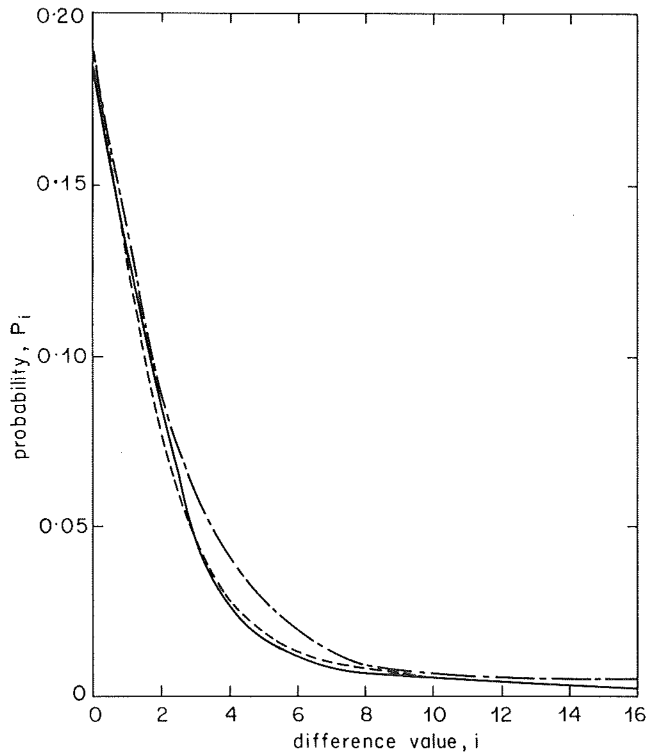
$$H(P^*) = H\left(\frac{1}{N} \sum_{k=1}^{N} P[k]\right)$$

Now, $f(x) = -x \log_2 x$ is a convex function[13] since $d^2f(x)/dx^2 \leqslant 0$. Therefore the entropy function,

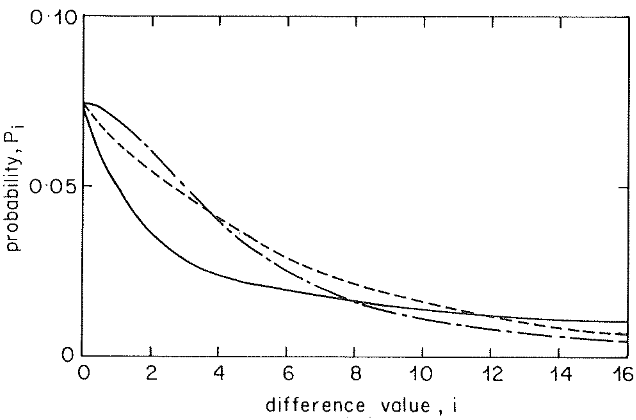
$H = -\sum_i p_i \log_2 p_i$, is also convex.



(a)



(b)



(c)

Fig. 8 - Measured probability density functions compared with two theoretical probability density functions

(a) Bowl of fruit (b) Girl with headscarf (c) Crowd scene

——— Measured —·— Laplace - - - - - Cauchy

A property of convex functions, $f(x)$ is that

$$\frac{1}{n} \sum_i f(x_i) \leqslant f\left(\frac{1}{n} \sum_i x_i\right)$$

for a series of values (x_i) of the variable x. Therefore $\hat{H} \leqslant H(P^*)$.

$\hat{H}$ is a lower bound on the bit-rate produced by an adaptive coder and $H(P^*)$ is a lower bound on the bit rate produced by a non-adaptive coder. Thus the expected length $E(\hat{L})$ of the bit stream from an adaptive coder is less than, or equal to, the expected length of the bit stream, for the same picture data, from a non-adaptive coder $E(L^*)$.

$$E(\hat{L}) \leqslant E(L^*)$$

This is also intuitively correct since a coder which always attempts to fit the code to the instantaneous properties of the signal will remove more redundancy than a coder which uses the average properties of the signals.

7.2. Probability-density functions for broadcast-quality digital video signals

O'Neal[29] and others have assumed and/or reported that d.p.c.m. difference samples have Laplacian probability-density functions. The present work differed from O'Neal's work in that the present work was concerned with exact differences for PAL video signals, as opposed to monochrome by O'Neal. It was found experimentally that the probability-density functions, for the video information, followed Cauchy distributions slightly more closely than Laplace distributions. This is shown by plotting the positive differences of Fig. 4, in Fig. 8, together with the corresponding Laplace and Cauchy distributions.

A Laplacian probability-density function is one where the value of the probability of each difference value i is given by the equation

$$p_i = \frac{1}{\sqrt{2}\sigma} \exp\, (\sqrt{2}|i|/\sigma)$$

As $1/\sqrt{2}\sigma$ is the probability of the difference value $i = 0$. σ can be found from the value of p_0.

In a Cauchy probability-density function

$$p_i = \frac{A}{A^2\pi^2 + i^2}$$

and $1/A\pi^2$ is the probability of difference value $i = 0$.

Laplace and Cauchy p.d.f.'s are defined by the probability of difference sample $i = 0$. Therefore, if the probability-density functions of exact difference samples are Cauchy distributions, these probability-density functions, and hence the entropy of the associated information sources, are defined by nothing more than the probability of difference sample $i = 0$.

Printed by BBC RESEARCH DEPARTMENT, Kingswood Warren, Tadworth, Surrey, KT20 6NP